Learning temporal coherence via self-supervision for GAN-based video generation

Year: 2020 | Citations: 253 | Authors: Mengyu Chu, You Xie, Jonas Mayer, L. Leal-Taix'e, Nils Thürey

---

Abstract

Our work explores temporal self-supervision for GAN-based video generation tasks. While adversarial training successfully yields generative models for a variety of areas, temporal relationships in the generated data are much less explored. Natural temporal changes are crucial for sequential generation tasks, e.g. video super-resolution and unpaired video translation. For the former, state-of-the-art methods often favor simpler norm losses such as L2 over adversarial training. However, their averaging nature easily leads to temporally smooth results with an undesirable lack of spatial detail. For unpaired video translation, existing approaches modify the generator networks to form spatio-temporal cycle consistencies. In contrast, we focus on improving learning objectives and propose a temporally self-supervised algorithm. For both tasks, we show that temporal adversarial learning is key to achieving temporally coherent solutions without sacrificing spatial detail. We also propose a novel Ping-Pong loss to improve the long-term temporal consistency. It effectively prevents recurrent networks from accumulating artifacts temporally without depressing detailed features. Additionally, we propose a first set of metrics to quantitatively evaluate the accuracy as well as the perceptual quality of the temporal evolution. A series of user studies confirm the rankings computed with these metrics. Code, data, models, and results are provided at https://github.com/thunil/TecoGAN.